Functional Gene Group Analysis Reveals a Role of Synaptic Heterotrimeric G Proteins in Cognitive Ability

Dina Ruano, Gonçalo R. Abecasis, Beate Glaser, Esther S. Lips, L. Niels Cornelisse, Arthur P.H. de Jong, David M. Evans, George Davey Smith, Nicolas J. Timpson, August B. Smit, Peter Heutink, Matthijs Verhage, and Danielle Posthuma

Although cognitive ability is a highly heritable trait, only a few genes have been identified, explaining relatively low proportions of the observed trait variation. This implies that hundreds of genes of small effect may be of importance for cognitive ability. We applied an innovative method in which we tested for the effect of groups of genes defined according to cellular function (functional gene group analysis). Using an initial sample of 627 subjects, this functional gene group analysis detected that synaptic heterotrimeric guanine nucleotide binding proteins (G proteins) play an important role in cognitive ability ($P_{EMP} = 1.9 \times 10^{-4}$). The association with heterotrimeric G proteins was validated in an independent population sample of 1507 subjects. Heterotrimeric G proteins are central relay factors between the activation of plasma membrane receptors by extracellular ligands and the cellular responses that these induce, and they can be considered a point of convergence, or a "signaling bottleneck." Although alterations in synaptic signaling processes may not be the exclusive explanation for the association of heterotrimeric G proteins with cognitive ability, such alterations may prominently affect the properties of neuronal networks in the brain in such a manner that impaired cognitive ability and lower intelligence are observed. The reported association of synaptic heterotrimeric G proteins with cognitive ability clearly points to a new direction in the study of the genetic basis of cognitive ability.

Introduction

Cognitive ability is a highly heritable trait, with heritability estimates ranging from 40% in young childhood to 80% in late adulthood. Disturbance in cognitive functioning ultimately causes lower test intelligence and is related to various psychiatric conditions, including schizophrenia, mental retardation, and autism. Identifying the actual genetic (and environmental) factors influencing cognitive ability may therefore elucidate the etiological basis of individual differences in both normal and abnormal cognitive functioning. Typically, each of the reported DNA variants associated with cognitive ability explain less than 1%–2% of the variation, and common variants in only a few genes (cholinergic receptor, muscarinic 2, CHRM2 [MIM 118493], catechol-O-methyltransferase, COMT [MIM 116790], and brain-derived neurotrophic factor, BDNF [MIM 113505]) have repeatedly shown associations. Despite the high heritability of cognitive ability and the major efforts made to understand it during the past decades, we are currently left with the generally accepted conclusion that “cognitive ability is most likely influenced by many genes of small effect, that possibly interact,” without knowing the identity of these genes. In the past decade, the scale of genotyping and genetic association studies has increased rapidly, from single-locus analysis to genome-wide association studies (GWAS) that allow the screening of 500,000–1,000,000 SNPs, covering 65%–95% of the human genome, depending on the array of choice. This has proven a successful method of identifying common genes of relatively large effect but has been less effective when rare variants of large effect are of importance or when identifying genes of small effect for complex traits, such as attention-deficit hyperactivity disorder (ADHD [MIM 143465]) or schizophrenia (MIM 181500).

Collective testing of genes involved in biological pathways has emerged as an alternative strategy for testing the combined effects of genetic variants with small effect size. Such pathways are usually defined as a set of proteins that participate in cascades of intracellular reactions, often triggered by extracellular ligands, involving enzyme-catalyzed posttranslational modifications of proteins and/or changes in their subcellular distribution. This ultimately leads to changes in cellular responses, such as altered gene expression or adaptation of cell morphology. Indeed, testing the combined effect of multiple genetic variants in such pathways has been shown to be more powerful than testing single-gene effects. Many proteins, however, are known to act across pathways, suggesting that they are not exclusively linked to one pathway. For instance, different neuromodulators, such as dopamine or 5HT, activate receptors in the plasma membrane, which are functionally and often structurally similar to the receptors in other
pathways. A higher degree of convergence exists for downstream signaling steps, such as guanine nucleotide binding proteins (G proteins), effector enzymes, second messengers, kinases, and classes of substrates. It is conceivable that genetic variation that influences complex traits accumulates at foci of convergence that act across different pathways—but have similar cellular function—and that can influence multiple biological pathways in a similar manner. We refer to grouping genes according to cellular function as “horizontal grouping,” as opposed to the classical grouping of genes in pathways, which we refer to as “vertical grouping” (Figure 1).

Horizontal grouping may be especially powerful in synaptic protein networks. Many different pathways have been described, all of which modulate synaptic function. Importantly, these pathways have a high degree of convergence, which has a common consequence; i.e., many modulate synaptic strength and change the coupling between neurons. Therefore, it is conceivable that genetic variation in (many) different pathways leads to similar consequences in synaptic function and hence to a common phenotypic effect. Collective testing of genes across pathways addresses this possibility.

Previous gene-finding studies for cognitive ability have focused on targeted candidate-gene testing or whole-genome linkage or associations studies. Here, we focus on evaluating the combined effect of multiple genes in both vertical pathways and horizontal functional gene groups. Results from the functional gene group approach clearly suggest the involvement of synaptic heterotrimeric G proteins in cognitive ability.

Subjects and Methods

Study Subjects
Subjects were part of the international collaborative research effort on the genetics of ADHD (the International Multi-Center ADHD Gene project [IMAGE]), and data of these subjects were obtained from the Database of Genotypes and Phenotypes (dbGaP). The IMAGE project was one of the six projects selected by the National Institutes of Health (NIH) in 2006 to be genotyped in the first phase of the Genetic Association Information Network (GAIN).

Characterization of the samples used in the IMAGE project has been described elsewhere in detail. In brief, 947 nuclear families of European descent (2844 individuals) from eight countries (Belgium, England, Germany, Holland, Ireland, Israel, Spain, and...
Switzerland) were included in the analysis. In this sample, standardized intelligence quotient (IQ) scores were available for 627 unrelated subjects (87.7% males). The age at the time of data collection for these 627 subjects ranged from 5 to 19 yrs, with a mean age of 11.0 (standard deviation [SD] = 2.7). Their IQ was assessed with the Wechsler Intelligence Scale for Children (WISC-III-R), in accordance with procedures described by SATTLER. No information on medication status or treatment was included in the analysis.

Genotyping and Quality Control
Genotyping was performed by Perlegen Sciences with the use of their genotyping platform, which comprises 600,000 tagging SNPs designed to be in high linkage disequilibrium with untyped SNPs for the three HapMap populations. Genotype data were cleaned by the National Center for Biotechnology Information (NCBI) with the use of the GAIN QA/QC software package before being uploaded to the dbGaP database. Details of the quality control (QC) process have been reported elsewhere. A total of 438,783 SNPs survived QC and were available in the downloaded data set.

Definition of Functional Gene Groups and Biological Pathways
As opposed to defining groups of genes according to the neuromodulator involved, we grouped genes horizontally; that is, across the classically defined pathways. Functional gene grouping was based on cellular function and relies on previous protein identification and data mining for synaptic genes and gene function. Because no selective purification method exists for the purification and analysis of presynaptic proteins, the criterion for proteins from the presynapse was based on manually curated data mining and confirmed by comparison to proteomic data from whole-synapse (pre plus post) analyses and previously characterized presynaptic subcomplexes in solubilized preparations.

Determination of genes expressed in the presynaptic terminal was based on the following set of criteria:

- Null mutation produces a secretion phenotype in synapses.
- Activation of the gene product (e.g., receptor) or blockade thereof modulates secretion in synapses.
- Immuno-electronmicroscopy detects the gene product in the terminal.
- The gene product is enriched in affinity-purified synaptic vesicle fraction.
- Overexpression produces a secretion phenotype in synapses.
- Immunocytochemistry colocalizes the gene product specifically with the presynaptic marker.
- The gene product is enriched in purified brain synaptosomes.
- Null mutation produces a secretion phenotype in secretory cells and the gene is expressed in the human brain.

Inclusion of postsynaptic proteins was based on the presence of proteins in proteomics analyses of the synapse, including preparations of synaptic membrane fractions as well as the postsynaptic density.

Synaptic genes are subdivided into 17 functional groups of genes on the basis of shared function into a biological process (i.e., horizontal grouping) and manually curated published data. These groups are as follows: cell adhesion and transsynaptic signaling molecules; cell metabolism (synaptic metabolic enzymes and their cofactors, excluding mitochondrial proteins); endocytosis (proteins involved in endocytosis); excitability (voltage-gated ion channels); exocytosis (proteins involved in regulated secretion); G protein relay (G protein subunits); GPCR signaling (G protein-coupled receptors); intracellular signal transduction (enzymes downstream of G protein or tyrosine kinase [TK] signaling); intracellular trafficking (vesicle adaptors, sorting proteins, motor proteins); ion balance/transport (ion and solute carriers and exchangers); ligand-gated ion-channel signaling; neurotransmitter metabolism (metabolizing enzymes); peptide/neurotrophic signaling (neuropeptide, trophic factors, hormones); protein clustering (scaffolding proteins); RNA and protein synthesis, folding, and breakdown; structural plasticity (cytoskeletal proteins and their regulators); and TK signaling (tyrosine receptor kinases).

One group was formed of remaining genes that are known to be expressed in the synapse but do not share any cellular function with other genes (this gene group was called "unknown").

In addition, four canonical pathways (vertical grouping) for synaptic modulation were defined: the dopaminergic, glutamatergic, serotonergic, and cannabinoid pathways. These pathway definitions are in line with previously published definitions and can be obtained, for example, from regular textbooks or online biological-pathway definitions. All locus IDs in each functional gene group and canonical pathway, as well as the main references used to determine the functional groups, are available online in Tables S1 and S2.

Data Processing
A schematic overview of the sequence of data-processing steps applied in this study is provided in Figure 2.

SNP Annotation
All SNPs that survived QC in the GAIN-IMAGE project were annotated according to the criteria of NCBI’s SNP database (dbSNP). All SNPs that occur within or in the vicinity of genes were identified. A SNP was assigned to a gene on the basis of positions of the gene’s reference sequence (RefSeq), with the addition of 2 kb to the 5’ end and 500 bp to the 3’ end of the gene. This region has been shown to include promoters (between −300 and −50 bp from the transcriptional start site [TSS]), as well as putative negative elements (between −1000 and −500 upstream of the TSS) and canonical hexanucleotide signals (20–50 bp from the 3’ end of the pre-RNA). SNPs mapped to multiple genes were assigned to each of these genes.

Statistical Analysis Method for Collective Testing of the Association of Gene Groups or Pathways
First, SNP-by-SNP analysis was carried out via linear regression. The \( \Sigma \log_{10}(P) \) was calculated with the use of all \( p \) values obtained by single SNP analysis of selected SNPs in a group of genes, as defined above. The \( \Sigma \log_{10}(P) \) is, however, not directly interpretable, because SNPs with the lowest \( p \) values will have the largest contribution to the calculated sum. As a result, this procedure can detect association of gene groups with a relatively low number of SNPs with low \( p \) values, even if some SNPs in the group have conventionally nonsignificant \( p \) values. In addition, the number of SNPs in a group will affect the test statistic \( \Sigma \log_{10}(P) \), in the sense that a group containing more SNPs will have a higher \( \Sigma \log_{10}(P) \) as compared to a group with fewer SNPs, assuming similar \( p \) values per SNP. Also, groups that contain SNPs with low \( p \) values and multiple SNPs in high linkage disequilibrium (LD) with those SNPs will have a higher \( \Sigma \log_{10}(P) \) than will groups
that contain SNPs of such low p values but without multiple SNPs in LD with those SNPs, given that p values are a function of LD between SNPs. Sample size will also affect the S-log10(P), because larger sample sizes are more likely to yield lower p values and thus higher S-log10(P).

So that unbiased interpretation of the test statistic was allowed, permutations (n = 10,000, or 100,000 when necessary) were carried out, conditional upon these factors, by permutation of phenotypes (i.e., the IQ scores) over genotypes. With this permutation procedure, only the relation between any genetic variant and the phenotype is disconnected, whereas the genomic haplotypic structure is kept intact. In addition, each selected group of genes will include exactly the same number of SNPs and genes, as well as the same haplotypic structure, and is based on the same sample size as that of the original data set.

For each group, we then determined by permutation how likely the observed S-log10(P) of that group was, given the number of SNPs and genes, the LD structure, and the sample size. A conservative Bonferroni correction was used to correct for multiple testing of different gene groups. Because we tested 23 different gene groups (all synaptic genes + 17 functional gene groups + one miscellaneous group and four biological pathways), the significance level was set at 0.05/23 = 0.0022. Given this significance level, the discovery sample of 627 subjects had sufficient power for detection of effect sizes of at least 2.4% of the total variation.

Software
All SNP-by-SNP analyses were conducted in PLINK. SNP selection, permutation of the data sets, and calculations of the combined effect were implemented in scripts written in the R software package.

Results
IQ Scores Normally Distributed in the GAIN-ADHD Sample
Phenotypic and genotypic data were obtained from dbGAP. IQ scores from 627 individuals were quantified with the Wechsler Intelligence Scales for Children (WISC) and had a mean of 100.7 (SD = 15.7) and a median of 101.6. Skewness was calculated to be 0.063, and the excess kurtosis relative to a normal distribution was −0.057 (Figure 3). Although this sample was originally ascertained for ADHD, this suggests that IQ scores are normally distributed in this sample.

Genome-wide SNP-by-SNP Analysis Does Not Detect Robust Associations
After QC, a total number of 438,783 SNPs were available for genome-wide association analysis. Of these, 179,725 SNPs mapped to 16,674 genes. The median number of SNPs per gene was 4.0, and the mean was 11.1, ranging from 1 to 1271 SNPs. First, we tested whether variation in single genes showed robust association with cognitive...
ability by applying a standard (SNP-by-SNP) genome-wide association analysis with the use of regression analysis implemented in PLINK and including only SNPs mapped to genes. Results of the SNP-by-SNP association analyses for all SNPs mapped to genes are shown in Figure 4.

With the genome-wide threshold being $1 \times 10^{-8}$, none of the single SNPs reached significance. Obviously, this demonstrates mainly a lack of power and underscores the need for larger sample sizes or more sophisticated use of available information. With the current sample size of 627 subjects and a genome-wide significance level of $1 \times 10^{-8}$, there was sufficient power for the detection of SNPs explaining at least 6.7% of the variance (Genetic Power Calculator). For the detection of SNPs of small effect size (e.g., 2%) with reasonable power (0.80), a sample of 2138 subjects would have been needed.

Analysis of Synaptic Functional Gene Groups Identifies Association with Cognitive Ability

Before testing detailed functional groups, we first tested whether genes expressed in synapses are associated with cognitive ability. This group contains 1024 genes, of which 900 (22,324 SNPs) were included on the Perlegen chip. Our results suggest that SNPs associated with cognitive ability are not randomly distributed across the genome, but tend to cluster in genes that are known to be expressed in synapses ($p = 0.001$) (Figure 5).

The 900 genes that are known to be expressed in the synapse and were included in the Perlegen chip were assigned to one of 17 functional gene groups on the basis of cellular function. Because some of the synaptic genes did not fit into any of the functional synaptic gene networks defined, we also tested this group of remaining synaptic genes as a negative control (the “unknown” group). The 49 genes in this group are expressed in the synaptic terminal, but their function is currently unknown.

Table 1 lists the results of the joint association analysis of SNPs within each of the functional groups, and Figure 6 and Figure 7 provide the quantile-quantile (Q-Q) plots for each of the tested functional gene groups.

As expected, the “unknown” group did not show any evidence for an association with cognitive ability (empirical $P_{EMP} = 0.880$). The group of synaptic heterotrimeric...
The American Journal of Human Genetics

Association with Cognitive Ability that was better than G proteins, however, yielded an overall test statistic for the sum of the negative log10 of the p value in the original data and the empirical permutations of the data. Bold values indicate significance after correction for multiple testing. \( P_{\text{EMP}} \) denotes empirical p value, based on 10,000 (or 100,000 when necessary) permutations of the data. The groups have been ordered by their p values. For each selected group, the number of genes and the number of SNPs are listed. In the last two columns, the sum of the negative log10 of the p value in the original data and the empirical p value are listed for each group.

G proteins, however, yielded an overall test statistic for association with cognitive ability that was better than that expected for a group that size on the basis of chance alone \( (P_{\text{EMP}} = 1.9 \times 10^{-5}) \), given the preset threshold of 0.0022. This suggests that the combined effect of genes within this group plays a role in explaining variation in cognitive ability. In the case of a single SNP, such a p value would correspond to an overall calculated effect size of at least 3.3% of the variation in cognitive ability. However, it should be noted that this calculated effect size (based on p value, sample size, and significance level) is based on single-SNP effects and is difficult to interpret when many (nonindependent) SNPs are involved.

None of the genes in the group of heterotrimeric G proteins have been associated with cognitive ability previously. This functional gene group consists of the following genes:


Two genes in this group were not included in our analyses because there were no SNPs available in the Perlegen chip: GNAI1 and GNAS. Notably, none of the single SNPs within this network would have been detected in the context of a genome-wide SNP-by-SNP analysis, with the lowest p value being \( 4.9 \times 10^{-5} \). In fact, only four SNPs within this functional gene group reached a p value below \( 10^{-3} \), and only 12.8% of SNPs had a p value below 0.05. Three of these four SNPs with a p value below \( 10^{-3} \) are in the GNAQ gene, and the 46 SNPs with a p value lower than 0.05 are distributed throughout 11 different genes: GNA14, GNAQ, GN2, GN5, GN4, GN11, GN5, GNA11, GN1, GN401, GN3, with frequencies of 13, 9, 8, 5, 3, 3, 1, 1, 1, and 1, respectively. This suggests that the effect of the functional group cannot be explained by the effect of a few individual SNPs or genes but must be ascribed to the combined effect of multiple genes in the functional gene group. This is also evident in the Q-Q plot of the heterotrimeric G proteins in Figure 6 (upper left panel), which shows that the distribution of the observed p values for the functional group of genes encoding heterotrimeric G proteins is consistently lower than
expected under the null hypothesis of uniform distribution, with no single p value standing out.

**Biological Pathway Analysis Shows No Association with Cognitive Ability**

Because the total collection of synaptic genes was associated with cognitive ability, we tested four biological pathways that are known to involve synaptic functioning as one of their roles in cognitive ability. These pathways involve the dopaminergic, glutamatergic, serotonergic, and cannabinoid pathways and follow the rationale of vertical grouping. Of these pathways, the serotonin pathway showed the largest number of genes overlapping with the synaptic genes. The dopaminergic, glutamatergic, serotonergic, and cannabinoid pathways yielded the following empirically derived p values: dopaminergic, $P_{EMP} = 0.5006$; glutamatergic, $P_{EMP} = 0.3883$; serotonergic, $P_{EMP} = 0.6211$; cannabinoid, $P_{EMP} = 0.8309$ (see also Table 1 and Figure 7), suggesting that collective testing of genes in synaptically relevant biological pathways is less successful in identifying genetic variation underlying cognitive ability than collectively testing genes that are grouped according to function in a biological process (horizontal grouping).
Correction for Population Stratification
The possible effects of population stratification on our results were explored via two methods. The Q-Q plot (corrected for \( \lambda \)) based on all SNPs available for analysis is provided in Figure 8A. Without correction for \( \lambda \), a slight deviation from the uncorrected expected distribution of p values under the null hypothesis was present, which was quantified in a genomic inflation factor of \( \lambda = 1.05672 \). This deviation is may be due to true association or may be indicative of false positives due to population stratification. Applying the genomic control correction method, we corrected all test statistics by the genomic inflation factor. Although all synaptic genes were no longer statistically significant as a group, this did not significantly change the results for the heterotrimeric G proteins (\( P_\text{EMP} = 0.00062 \)).

Second, given that the primary sample is known to include subsamples as a result of data being collected in different sites and countries, we calculated Z scores within each site and conducted all analyses again. The Z score
On the basis of these results, we conclude that possible spurious effects due to population stratification cannot account for the detected association of the heterotrimeric G proteins and cognitive ability.

**Validation of Significant Functional Gene Group in Independent Population-Based Sample**

The Avon Longitudinal Study of Parents and Children (ALSPAC) study served as a validation sample. ALSPAC is a UK-based, population-based, prospective birth cohort with extensive data collection on health and development of children and their parents, predominantly those of white European origin, and has been described previously. Ethical approval for the study was obtained from the ALSPAC Law and Ethics Committee and the local research ethics committees. Genotyping on 1543 individuals was initially performed with the Illumina HumanHap 300K BeadChip for 1568 blood-derived DNA samples. After QC, the clean data set comprised 1507 samples (excluding individuals with potential non-European ancestry, more than 5% missing genotype data, sex-inconsistent X-heterozygosity or a genome-wide heterozygosity of more than 36.4% or less than 34.3%).

Because this validation sample is based on a general population sample and does not consist of individuals ascertained on the basis of ADHD, any replicated effect would confirm that the observed association is related to cognitive ability in general and is not specific to individual differences in cognitive ability in an ADHD population. The validation sample (n = 1507) had 100% power for detection of an effect size of 3.3% against a significance level of 0.05 (one test conducted).

Cognitive ability was measured in children 8 yrs of age with the Wechsler Intelligence Scale for Children (WISC-III). A short version of the test, consisting of alternate items only (with the exception of the coding task), was conducted by trained psychologists. Verbal (information, similarities, arithmetic, vocabulary, comprehension) and performance (picture completion, coding, picture arrangement, block design, object assembly) subscales were administered, the subtests scaled and scores for total IQ derived.

Of the 27 genes in the heterotrimeric G protein relay group, four genes were not covered in the validation sample. The validation sample included 265 SNPs mapped to 23 genes (versus 359 SNPs mapped to 25 genes in the original sample) in the G protein group. The difference in available SNPs was due to the difference in platforms used in the original and validation samples. The two genes (GNB2 and GNG11) that were present in the original sample but were not covered in the validation sample include the GNG11 gene, which was one of the most significant genes of the group, with three SNPs showing a p value < 0.05 (rs4262 = 0.009793; rs180236 = 0.01612; rs180241 = 0.02378). Figure 9 shows the Q-Q plot of all tested SNPs in the validation sample and suggests that all or most SNPs are important and that no single SNP drives the observed association of the G protein group.
An r² of 1. For reasons of comparison, gene coverage was between the two data sets and another seven SNPs have 10

The American Journal of Human Genetics

Here, we applied an innovative functional gene group approach to cognitive ability. Functional gene groups were defined on the basis of shared cellular function of genes, as determined by previous protein identification and data mining for synaptic genes and gene function. Initially, we identified the group of synaptic heterotrimeric G proteins to be associated with cognitive ability in a relatively small sample of 627 subjects. We replicated these findings in an independent large cohort of 1507 subjects.

Association of the group of heterotrimeric G proteins could not be attributed to a single gene or a single SNP, because none of the individual p values came close to genome-wide significance, confirming the importance of focusing on the gene group as the unit of analysis and not on single-SNP effects.

Heterotrimeric G proteins consist of three subunits (α, β, and γ), for which a total of 33 genes are found in the human genome. Twenty-seven of these are ubiquitously expressed in the synapse. Many of these proteins are also expressed in the nonsynaptic areas of neurons and in other cells inside and outside the brain. Heterotrimeric G proteins are central relay factors between the activation of plasma membrane receptors by extracellular ligands and the cellular responses that these induce. Some signaling molecules in the brain can also activate ionic currents upon binding to ionotropic receptors, but for most signaling molecules no parallel or alternative pathways next to G protein receptor-coupled signaling exist to induce cellular responses. Therefore, heterotrimeric G proteins may be considered a point of convergence; a kind of “signaling bottleneck.” Although alterations in synaptic processes may not be the exclusive explanation for the association of heterotrimeric G proteins with cognitive ability, it is plausible that such alterations prominently affect the properties of neuronal networks in the brain in such a manner that impaired cognition and lower intelligence is observed. Synaptic processes are thought to have a central role in the “real time” processing capacities of the brain—for instance, in discrimination tasks, working memory, attention, and decision making—as well as in adaptations required for “long-term synaptic modulation” in learning and memory.

It is worth noting that the second most significant (although not below the conservative threshold of significance) functional gene group associated with cognitive ability in this study was the group of “transmitter synthesizing and metabolizing proteins,” which includes the COMT and monoamine oxidase A (MAOA [MIM 309850]) genes. Previous studies have systematically pointed to a role of metabolic enzymes in cognitive ability. In fact, the COMT gene has been associated with many different cognitive traits. These results suggest that further exploration of a role of the group of transmitter synthesizing and metabolizing proteins may be indicated.

Evaluating the Combined Effect of Multiple Genes in Functional Gene Groups

The functional gene group approach may also prove fruitful for other complex traits or common disorders that are

Figure 9. Q-Q Plot of All SNPs in the Heterotrimeric G Protein Relay Group in the Validation Sample

Of the 46 SNPs in the G protein group that had a p value < 0.05 in the original data set, 27 SNPs had a proxy SNP with an r² > 0.8, of which seven SNPs are identical between the two data sets and another seven SNPs have an r² of 1. For reasons of comparison, gene coverage was determined on the basis of LD structure and genomic density and was based on the HapMap CEU LD structure. It was calculated by the sum of the typed SNPs as well as the tagged SNPs divided by the total known common SNPs within a gene (see Table 2).

In line with the recommendation by Holmans et al., the validation study focused on the functional gene group rather than on a direct SNP-by-SNP replication. This may result in a conservative p value in the validation sample, given that the SNPs that showed most evidence for association in the initial sample are not directly included in the validation study. However, the main goal is to validate the association with the functional gene group, not with single SNPs. This is also in line with the assumption advocated here and in Holmans et al., that it is more powerful to treat the functional gene group as the unit of analysis and not the single SNPs.

Analyses were conducted in PLINK similarly to the method applied in the original sample, and 10,000 permutations were used to determine the empirical p value of the combined effect of all included SNPs. The Σ-log₁₀(P) of the heterotrimeric G proteins was 136, with an empirical p value of 0.047, validating our findings in an independent cohort.

Discussion

Here, we applied an innovative functional gene group approach to cognitive ability. Functional gene groups...
potentially influenced by many genes of small effect. One of the main assumptions of current GWAS is that common diseases are caused by at least a few common genetic variants of relatively large effect (the common disease, common variant [CDCV] hypothesis). If this assumption is not met (e.g., many common alleles with small effect, genetic heterogeneity), GWAS will not work, because in different individuals different variants will account for a disease status or trait level.38 However, if these different variants share a common molecular function, thus sustaining a common biological process, focusing on their combined effect will still be a valuable approach. The functional gene group approach may thus be able to detect genes even when there is large genetic heterogeneity at the SNP level, as long as the many alleles of small effect share a common function in a biological process.

Unlike previous methods, the functional gene group approach does not rely on post hoc formulation of pathways or networks, but instead takes a hypothesis-driven approach by directly testing functional gene groups. This is opposed to the commonly followed strategy in which the most significant SNPs from a genome-wide association analysis are annotated to search for possible biological pathways associated with the trait. Although this may

<table>
<thead>
<tr>
<th>Gene</th>
<th>LID</th>
<th>Total (N SNPs)</th>
<th>Initial Sample (N SNPs)</th>
<th>Validation Sample (N SNPs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Genotyped</td>
<td>Tagged&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>GNA11</td>
<td>2767</td>
<td>29 24</td>
<td>3 0</td>
<td>0.13</td>
</tr>
<tr>
<td>GNA12</td>
<td>2768</td>
<td>180 118</td>
<td>21 87</td>
<td>0.92</td>
</tr>
<tr>
<td>GNA13</td>
<td>10672</td>
<td>57 8</td>
<td>2 5</td>
<td>0.88</td>
</tr>
<tr>
<td>GNA14</td>
<td>9630</td>
<td>348 217</td>
<td>73 68</td>
<td>0.65</td>
</tr>
<tr>
<td>GNA15</td>
<td>2769</td>
<td>29 20</td>
<td>7 3</td>
<td>0.50</td>
</tr>
<tr>
<td>GNAI1</td>
<td>2770</td>
<td>138 71</td>
<td>19 23</td>
<td>0.59</td>
</tr>
<tr>
<td>GNAI2</td>
<td>2771</td>
<td>14 3</td>
<td>1 0</td>
<td>0.33</td>
</tr>
<tr>
<td>GNAI3</td>
<td>2772</td>
<td>61 26</td>
<td>4 16</td>
<td>0.77</td>
</tr>
<tr>
<td>GNAI4</td>
<td>2773</td>
<td>176 87</td>
<td>30 37</td>
<td>0.77</td>
</tr>
<tr>
<td>GNAO1</td>
<td>2775</td>
<td>242 143</td>
<td>29 60</td>
<td>0.62</td>
</tr>
<tr>
<td>GNAQ</td>
<td>2776</td>
<td>369 208</td>
<td>26 170</td>
<td>0.94</td>
</tr>
<tr>
<td>GNAS</td>
<td>2777</td>
<td>97 45</td>
<td>14 8</td>
<td>0.49</td>
</tr>
<tr>
<td>GNAI1</td>
<td>2778</td>
<td>3 1</td>
<td>0 0</td>
<td>0.00</td>
</tr>
<tr>
<td>GNAZ</td>
<td>2781</td>
<td>83 41</td>
<td>5 32</td>
<td>0.90</td>
</tr>
<tr>
<td>GNB1</td>
<td>2782</td>
<td>76 27</td>
<td>7 11</td>
<td>0.67</td>
</tr>
<tr>
<td>GNB2</td>
<td>2783</td>
<td>1 0</td>
<td>1 0</td>
<td>-</td>
</tr>
<tr>
<td>GNB3</td>
<td>2784</td>
<td>10 4</td>
<td>4 2</td>
<td>1.50</td>
</tr>
<tr>
<td>GNB4</td>
<td>59345</td>
<td>62 31</td>
<td>3 11</td>
<td>0.45</td>
</tr>
<tr>
<td>GNB5</td>
<td>10681</td>
<td>94 41</td>
<td>7 28</td>
<td>0.85</td>
</tr>
<tr>
<td>GNG10</td>
<td>2790</td>
<td>6 4</td>
<td>2 1</td>
<td>0.75</td>
</tr>
<tr>
<td>GNG11</td>
<td>2791</td>
<td>11 7</td>
<td>6 1</td>
<td>1.00</td>
</tr>
<tr>
<td>GNG12</td>
<td>55970</td>
<td>201 117</td>
<td>20 81</td>
<td>0.86</td>
</tr>
<tr>
<td>GNG2</td>
<td>54331</td>
<td>255 114</td>
<td>31 55</td>
<td>0.75</td>
</tr>
<tr>
<td>GNG3</td>
<td>2785</td>
<td>0 0</td>
<td>0 0</td>
<td>0.00</td>
</tr>
<tr>
<td>GNG4</td>
<td>2786</td>
<td>113 70</td>
<td>18 22</td>
<td>0.57</td>
</tr>
<tr>
<td>GNG5</td>
<td>2787</td>
<td>9 5</td>
<td>1 0</td>
<td>0.20</td>
</tr>
<tr>
<td>GNG7</td>
<td>2788</td>
<td>148 97</td>
<td>25 25</td>
<td>0.52</td>
</tr>
<tr>
<td>Total</td>
<td>2812</td>
<td>1529 359</td>
<td>746 264</td>
<td>0.72&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Abbreviations are as follows: LID, locus ID; Cov. Rate, coverage rate.

<sup>a</sup> Common SNPs, based on MAF > 0.05.
<sup>b</sup> Tagged SNPs, based on pairwise r² ≥ 0.80.
<sup>c</sup> Average coverage rate.
prove a successful strategy, it will not detect any pathways or functional networks of genes in which most genes are of small and equal effect size, because it focuses initially on SNPs with the largest effects. The applied permutation procedure renders the functional gene group analysis independent of the number of SNPs per gene, of the number of genes per pathway or functional group, and of gene differences in LD structure. Functional gene group analysis may be further preferred because it circumvents the multiple-testing problem and because effect sizes are likely to be increased in comparison to single-SNP effects (because these are now a function of the combined effect, rather than of effects of single genes). This approach may therefore aid in resolving the “case of the missing heritability.” It should be noted, however, that the effect size as estimated in the current study explained 3.3% of the observed variation in cognitive ability, which is considered large in comparison to effects of single genes for cognitive ability but is still modest in terms of estimated heritability of cognitive ability. Additional pathways or functional gene groups are therefore also likely to contribute to variation in cognitive ability.

To our knowledge, this is the first study reporting on a functional role of synaptic heterotrimeric G proteins in cognitive ability, and it thereby directs future research into the genetic basis of cognitive ability toward synaptic signaling processes. At the same time, these results underscore the notion that pathway analysis or group analysis is more informative as the unit of analysis than is single-SNP analysis, because it is directly related to biological function. The functional gene group approach adapted in this study relies on grouping genes according to similar cellular function on the basis of extensive lab experiments (wholesynapse analyses and solubilized preparations) and data mining. This functional gene group approach complements existing methods and may also be useful in identifying etiological factors underlying complex diseases for which classic, genome-wide, SNP-by-SNP analysis has been unsuccessful so far.

Supplemental Data

Supplemental Data include two tables and can be found with this article online at http://www.ajhg.org.

Acknowledgments

The work of D.R. is supported by the Portuguese Foundation for Science and Technology under grant no. SFRH / BPD / 28725 / 2006. D.P. is supported by NWO/VIDI 016-065-318. We would like to thank the Young Academy of the Royal Dutch Academy of Arts and Science and the Simonsfonds for providing funds to D.P. and D.R., respectively, to facilitate collaboration with G.R.A. A.B.S. is supported by the Centre for Medical Systems Biology (CMSB). Statistical analyses were carried out on the Genetic Cluster Computer, which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003). The genotyping of the samples was provided through the Genetic Association Information Network (GAIN). The data set(s) used for the analyses described in this manuscript were obtained from the dbGaP database through dbGaP accession no. phs000016.v1.p1. Samples and associated phenotype data for the Whole-Genome Association Study of Attention Deficit Hyperactivity Disorder were provided by S Faraone. We are extremely grateful to all the families who took part in this study, to the midwives for their help in recruiting them, and to the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. The UK Medical Research Council (grant ref. 74882), the Wellcome Trust (grant ref. 076467), and the University of Bristol provided core support for ALSPAC. D.P. wishes to thank Zoltan Bochdanovits and Peter Visscher for reading earlier versions of this paper and/or for valuable discussions on the applied methods. We thank Synaptologics BV (http://www.synaptologics.com) for providing the database of functional gene groups.

Received: September 9, 2009
Revised: November 26, 2009
Accepted: December 4, 2009
Published online: January 7, 2010

Web Resources

The URLs for data presented herein are as follows:
GAIN QA and QC software package http://www.sph.umich.edu/ csg/abecasis/GainQC/
PLINK software, http://pngu.mgh.harvard.edu/~purcell/plink/
R software, http://www.r-project.org

References


